XXIV General Assembly of European Geophysical Society
The Hague, 19-23 April 1999
HSA9.01 Open session on statistical methods in hydrology

An advanced method for preserving skewness in single-variate, multivariate, and disaggregation models in stochastic hydrology

Demetris Koutsoyiannis
Department of Water Resources
National Technical University of Athens

Introduction

$$Y_1 = b_{11} V_1$$
$$Y_2 = b_{21} V_1 + b_{22} V_2$$
$$Y_3 = b_{31} V_1 + b_{32} V_2 + b_{33} V_3$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

A simple stochastic model that generates hydrological variables Y_i using innovations (noise variables) V_i (with $\mathrm{Var}[V_i] = 1$)

Due to the central limit theorem:
The distribution of Y_i tends to be more symmetric for increasing number of innovations

For constant $\mu_3[V_i]$:
$|\mu_3[Y_1]| > |\mu_3[Y_2]| > |\mu_3[Y_3]| > \ldots$

Conversely:
To obtain a constant $\mu_3[Y_i]$ for increasing number of innovations, more and more skewed innovations V_i are needed:
$|\mu_3[V_1]| < |\mu_3[V_2]| < |\mu_3[V_3]| < \ldots$

In finite samples the skewness cannot be arbitrarily high
Theoretically the limit equals √(sample size)
Practically, a $|\mu_3[V_i]| > 0.5$ √(sample size) cannot be preserved and therefore $\mu_3[Y_i]$ is not preserved (Todini, 1980)

Mathematical framework

- **Model**:

  $\mathbf{Y} = \mathbf{a}\,\mathbf{Z} + \mathbf{b}\,\mathbf{V}$ ……………………

  where

  $\mathbf{Y}$: vector of variables to be generated

  $\mathbf{Z}$: vector of variables with known values

  $\mathbf{V}$: vector of innovations (with $\mathrm{Var}[V_i] = 1$)

  $\mathbf{a}$ and $\mathbf{b}$: matrices of parameters ($\mathbf{b}$ square)

- **Main parameter estimators**:

  $\mathbf{b}\,\mathbf{b}^T = \mathbf{c}$ ……………………………

  $\boldsymbol{\xi} := \mu_3[\mathbf{V}] = [\mathbf{b}^{(3)}]^{-1}\ \{\mu_3[\mathbf{Y}] - \mu_3[\mathbf{a}\,\mathbf{Z}]\}$…

  where

  $\mathbf{c} := \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}] - \mathbf{a}\,\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]\,\mathbf{a}^T$

  (equivalently, $\mathbf{c} := \mathrm{Cov}[\mathbf{Y} - \mathbf{a}\,\mathbf{Z}, \mathbf{Y} - \mathbf{a}\,\mathbf{Z}]$)

  $\mathbf{b}^{(3)}$: matrix with elements the cubes of $\mathbf{b}$

> - Representative for most common stochastic models in hydrology

> - Infinite solutions if **c** is positive definite
> - No solutions otherwise (inconsistent **c**)
> - There exist two algorithms for determining (different solutions) **b**
>   - Cholesky decomposition (triangular **b**)
>   - Singular value decomposition (based on eigenvectors of **b**)

> - The skewness of **V** depends on **b**
> - If some element of $\boldsymbol{\xi} = \mu_3[\mathbf{V}]$ is too high then $\mu_3[\mathbf{Y}]$ will be not preserved

Problem formulation

Determine **b** from the known $\mathbf{c} = \mathbf{b}\,\mathbf{b}^T$ so that the coefficients of skewness of **V** be as small as possible

- For **c** positive definite:

  Find the optimal solution **b**, leading to the smallest value of $\max_i\{\xi_i\}$

  $\Rightarrow$ Optimisation problem (single-objective, unconstrained)

- For **c** not positive definite:

  Find a "solution" **b**, leading to a small departure of $\mathbf{b}\,\mathbf{b}^T$ from **c**, and simultaneously a small value of $\max_i\{\xi_i\}$

  $\Rightarrow$ Optimisation problem (multiple-objective, or single-objective constrained)

Example A: Temporal rainfall disaggregation

- Consider the generation of a rainfall event with duration $D = 20$ h using a half-hour time resolution ($k = 40$ half-hour rainfall increments Y_i, $i = 1, \ldots, 40$)
- Assume covariance structure of Y_i as in the *Scaling Model of Storm Hyetograph* (Koutsoyiannis and Foufoula-Georgiou, *Water Resources Research*, 29(7), 1993)

$$\mathrm{Cov}[Y_i, Y_j] = [(c_2 + c_1^2)\, f(|j - i|, \beta)\, k^{\beta} - c_1^2]\,(D^{2(\kappa + 1)} / k^2)$$

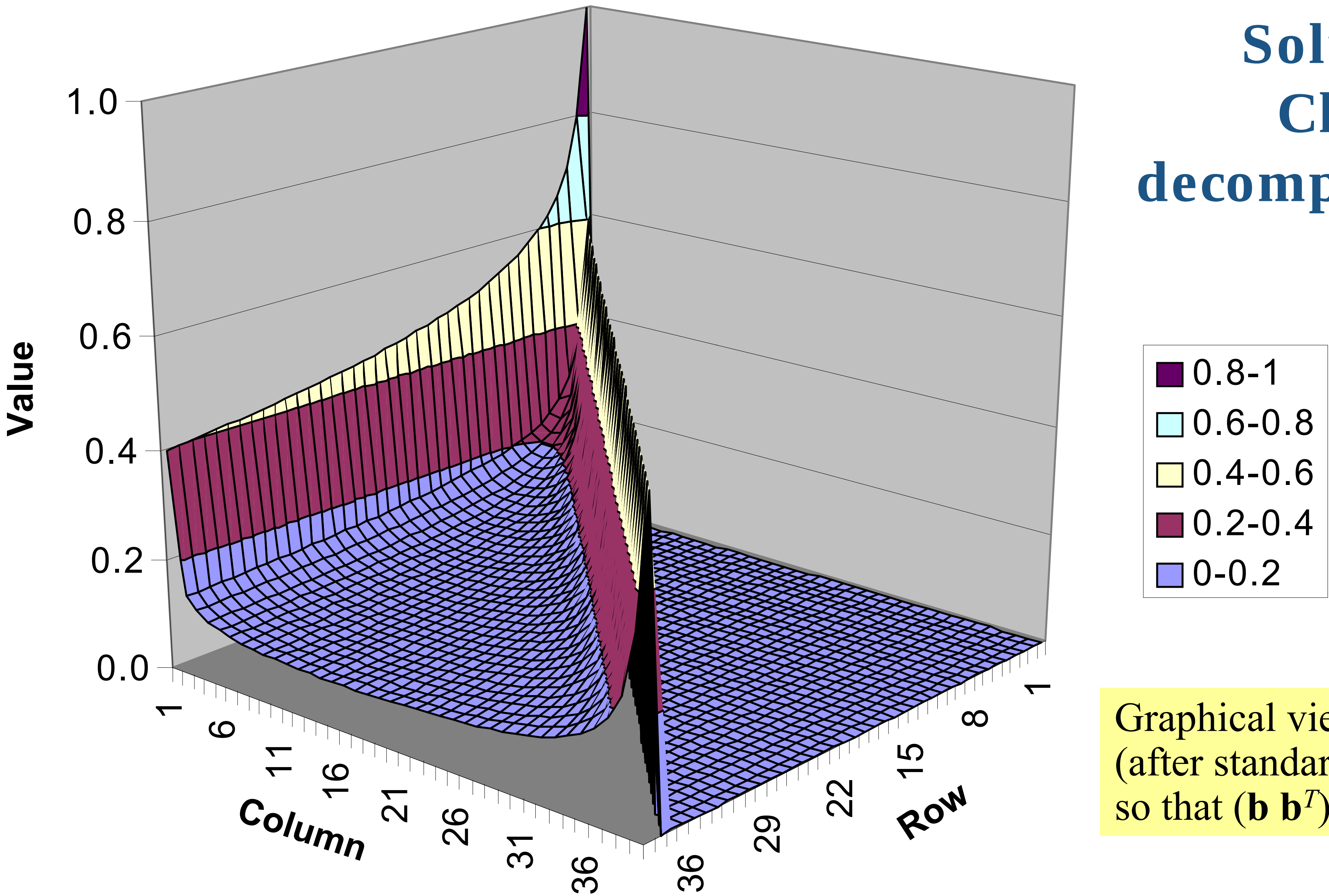
where

$$f(m, \beta) = (1/2)\,[(m - 1)^{2 - \beta} + (m + 1)^{2 - \beta}] - m^{2 - \beta} \quad \text{if } m > 0$$

$$f(m, \beta) = 1 \text{ if } m = 0$$

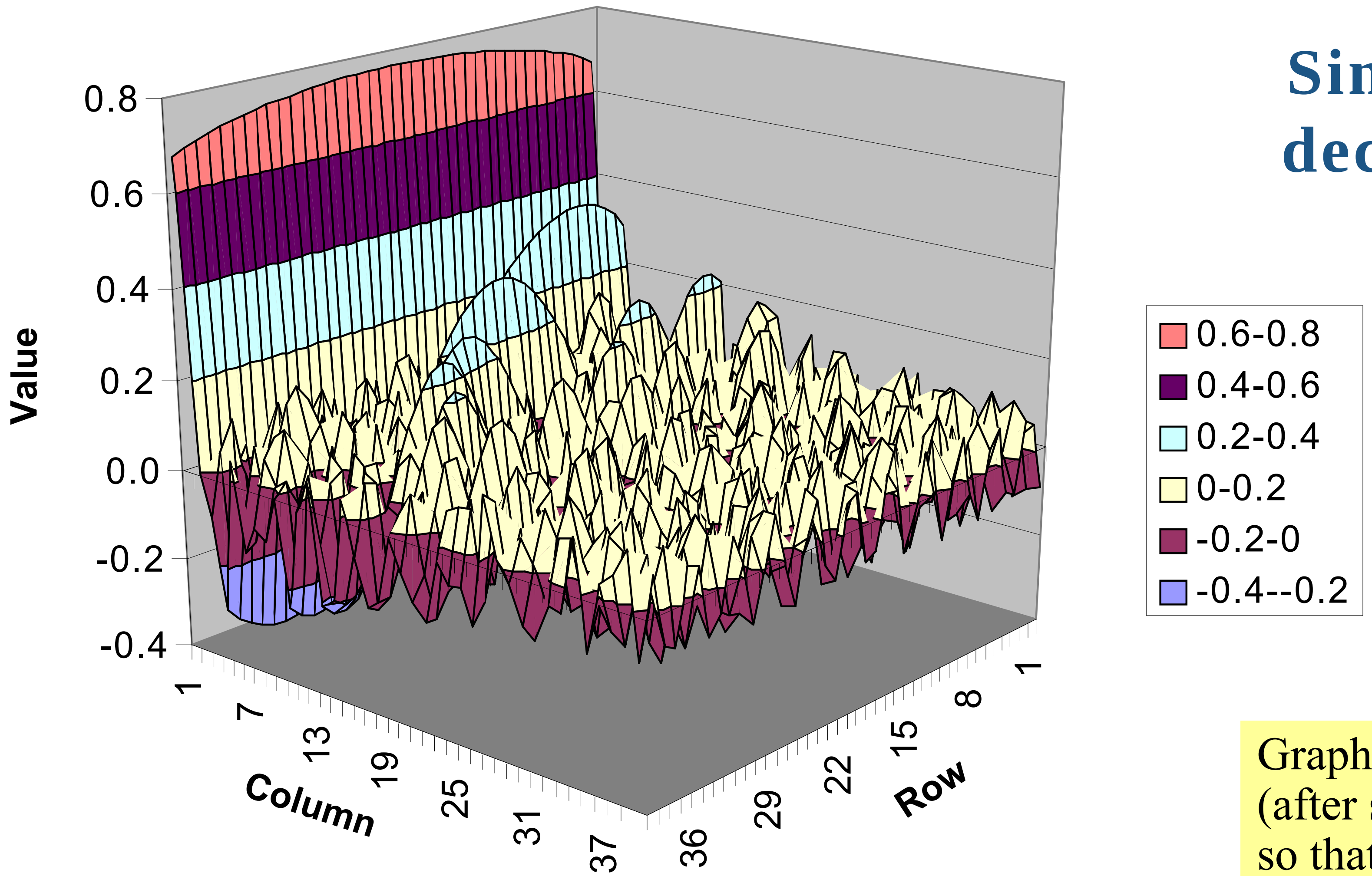
- Assume two parameter gamma distribution for Y_i
- Parameters: $c_1 = 8.74$, $c_2 = 85.68$, $\kappa = -0.449$, $\beta = 0.1$
- Statistics of Y_i: $E[Y_i] = 1.14$ mm, $C_\mathrm{v}[X_i] = 1.44$, $C_\mathrm{s}[X_i] = 2.88$
- Single variate problem with long memory (not a typical ARMA model)
- Generation model $\mathbf{Y} = \mathbf{b}\,\mathbf{V}$ with $\mathbf{b}\,\mathbf{b}^T = \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}]$
- $\mathbf{b}$ is a matrix of parameters with size 40×40 (1600 unknowns)

Example A – Solution 1: Cholesky decomposition



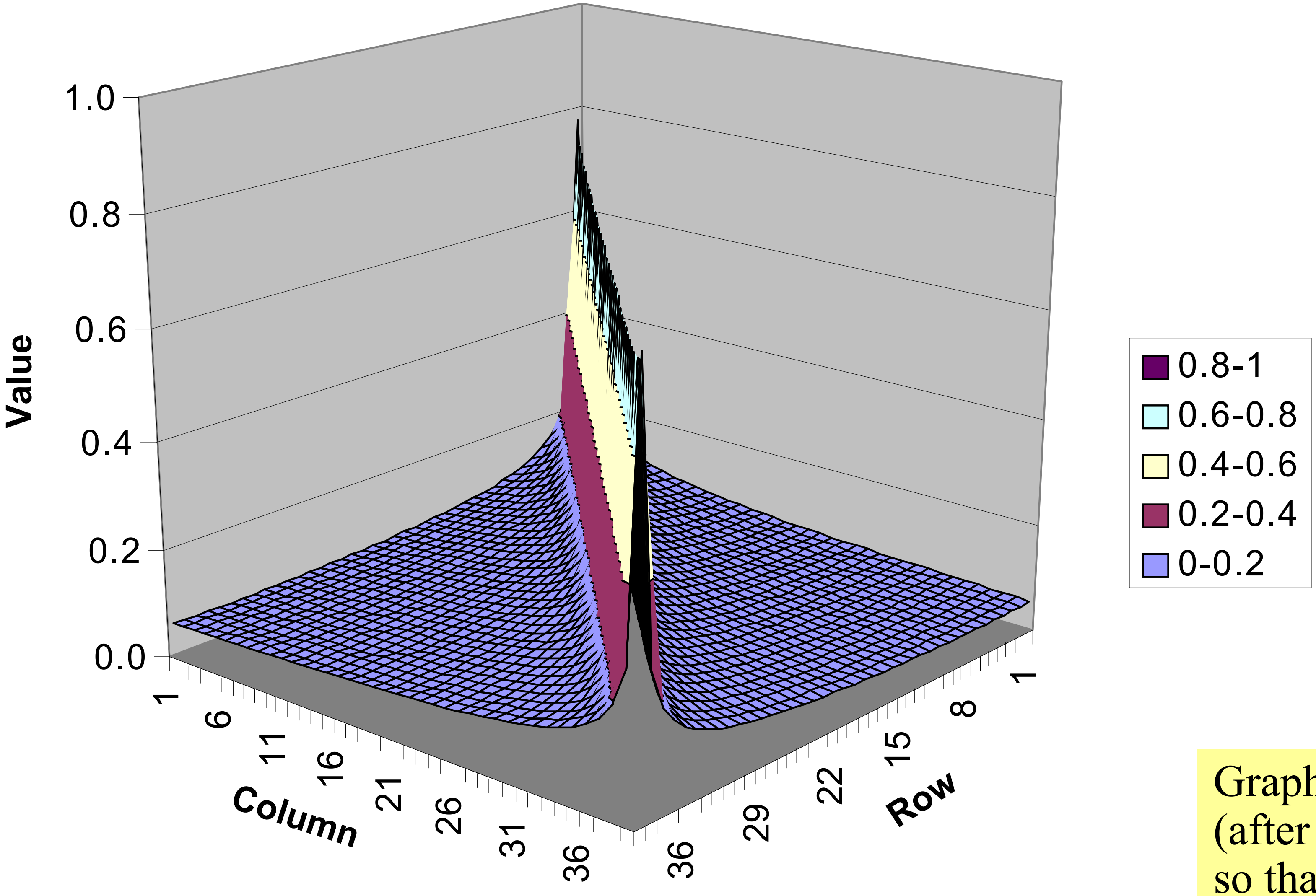
Graphical view of **b** (after standardisation so that $(\mathbf{b}\,\mathbf{b}^T)_{ii} = 1$)

Example A – Solution 2: Singular value decomposition



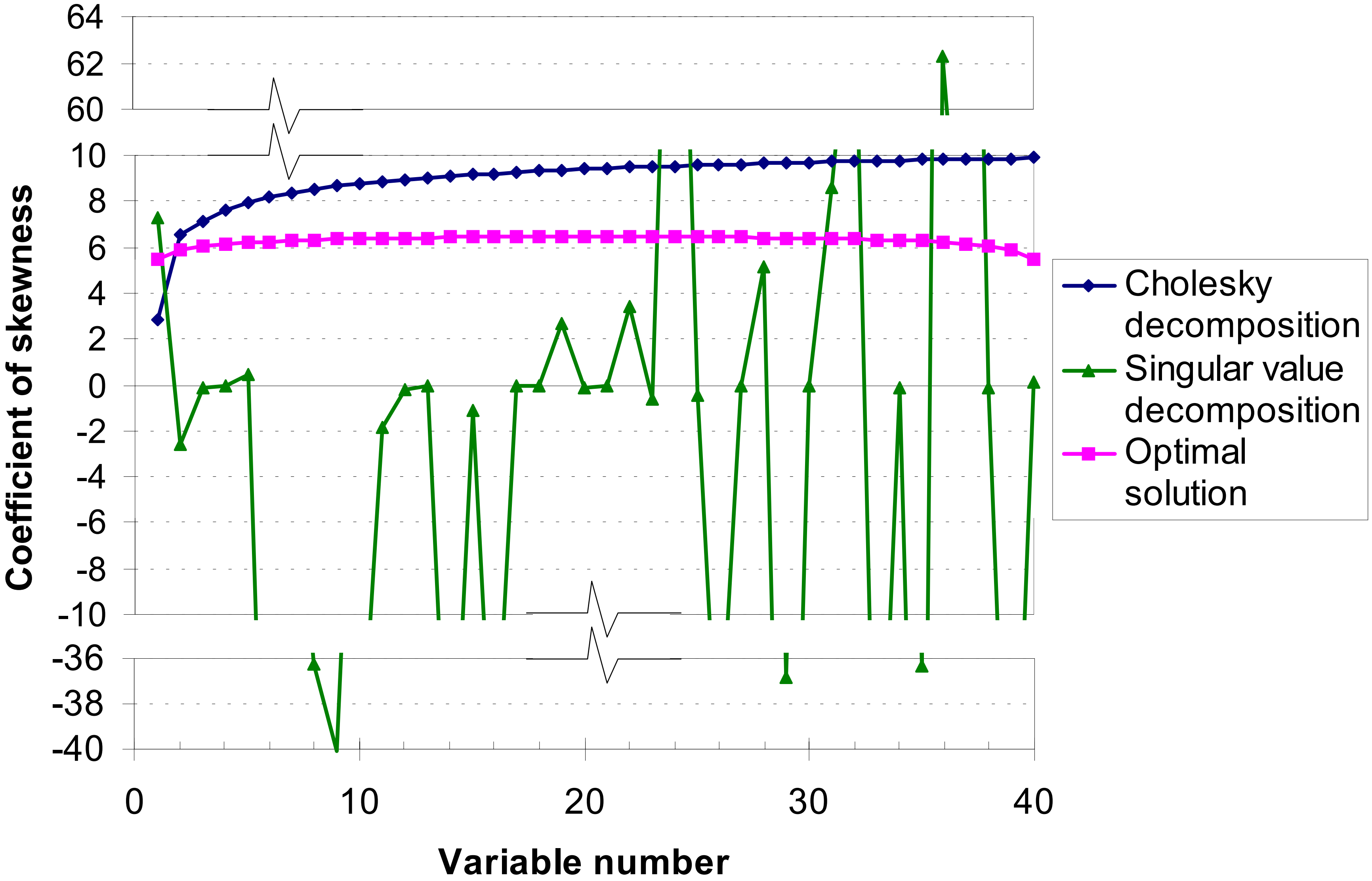
Graphical view of **b** (after standardisation so that $(\mathbf{b}\,\mathbf{b}^T)_{ii} = 1$)

Example A – Solution 3: Optimal



Graphical view of **b** (after standardisation so that $(\mathbf{b}\,\mathbf{b}^T)_{ii} = 1$)

Example A – Resulting coefficients of skewness for innovation variables



Example B: Multivariate generation of monthly rainfall and runoff

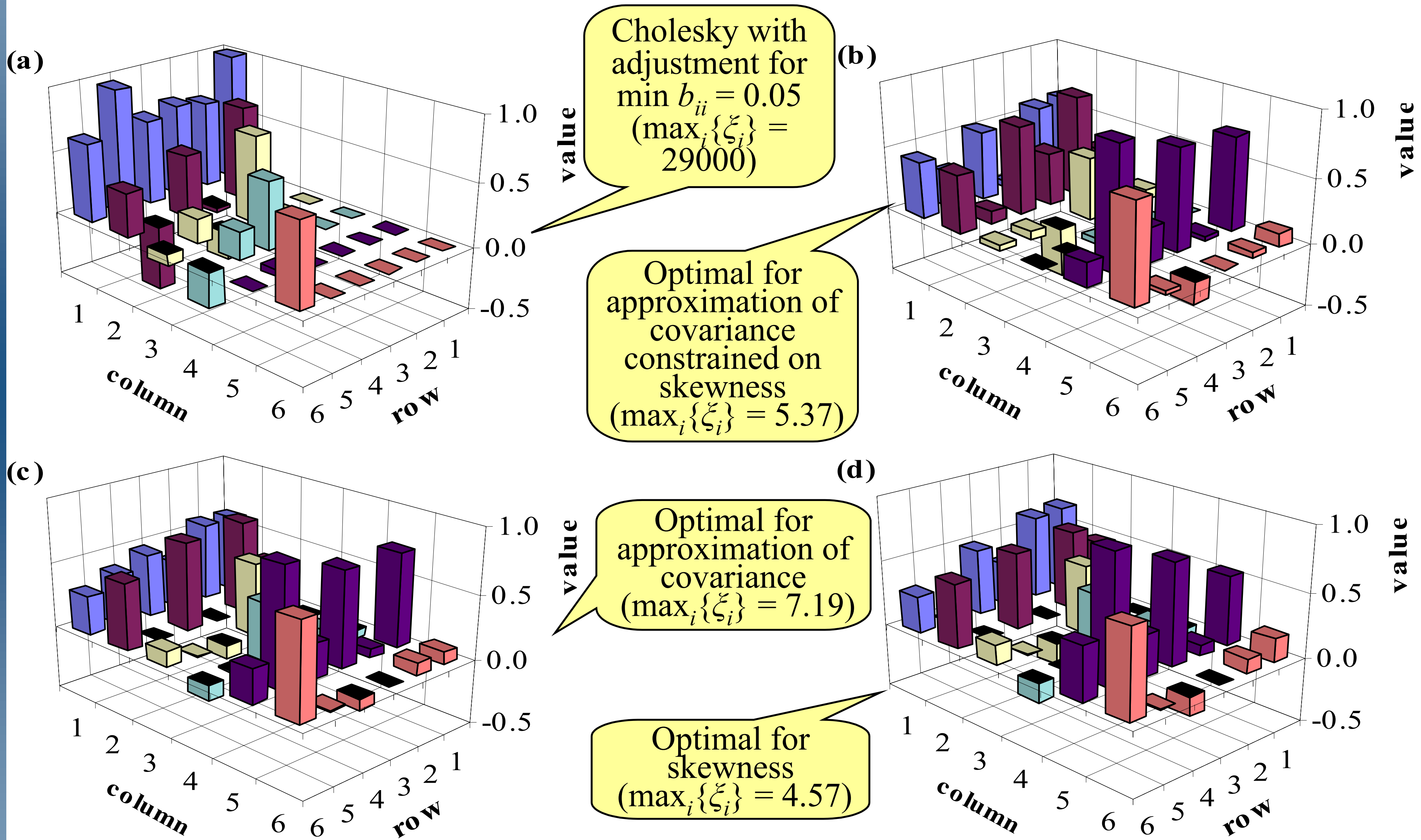
- Multivariate generation problem with 6 locations:
  - 2 variables: simultaneous monthly rainfall and runoff
  - 3 basins: Evinos, Mornos and Yliki, supplying water to Athens, Greece
- Model PAR(1):

$$\mathbf{Y} = \mathbf{a}\,\mathbf{Z} + \mathbf{b}\,\mathbf{V}$$

where $\mathbf{Y} \equiv \mathbf{X}^{s}$, $\mathbf{Z} \equiv \mathbf{X}^{s-1}$($s$ stands for subperiod, i.e., month; here $s = 8 \rightarrow$ May)

- Characteristic statistics:
  - Coefficients of skewness of Y_i: 0.76-1.49
  - Cross-correlation coefficients: 0.16-0.90
  - Autocorrelation coefficients of runoff: 0.60-0.80
  - Autocorrelation coefficients of rainfall: ≈ 0
  - Matrix $\mathbf{c} = \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}] - \mathbf{a}\,\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]\,\mathbf{a}^T$ is inconsistent (not positive definite)

Example B – Different solutions of matrix b



Proposed algorithm: Objective function

- Component 1: Preservation (or approximation) of covariances

  $||\mathbf{d}||^2 := \sum_i \sum_j d_{ij}^2$ where $\mathbf{d} := \mathbf{b}\,\mathbf{b}^T - \mathbf{c}$

- Component 2: Preservation of variances

  $||\mathbf{d^*}||^2 := \sum_i d_{ii}^2$ where $\mathbf{d}^* := \mathrm{diag}(d_{11}, \ldots, d_{nn})$

- Component 3: Preservation of skewness

  $||\boldsymbol{\xi}||_p^2 := \left(\sum_i |\xi_i|^p\right)^{2/p}$ where p a large integer so that $||\boldsymbol{\xi}||_p \approx \max_i\{|\xi_i|\}$

- Combination of the three components and problem solution by minimising

  $$\theta^2(\mathbf{b}) := (\lambda_1 / n^2)\, ||\mathbf{d}(\mathbf{b})||^2 + (\lambda_2 / n)\, ||\mathbf{d}^*(\mathbf{b})||^2 + \lambda_3\, ||\boldsymbol{\xi}(\mathbf{b})||$$

  where n is the matrix size, and λ_1, λ_2 and λ_3 adjustable multipliers

  typical values: $\lambda_1 = 1$, $\lambda_2 = 10^3$, $\lambda_3 = 10^{-3}$

Proposed algorithm: Optimisation procedure

- The matrix of derivatives of θ^2 with respect to the unknown parameters b'_{ij} has a very simple expression, i.e.,

  $$d\theta^2 / d\mathbf{b} = (4\,\lambda_1 / n^2)\;\mathbf{d}\,\mathbf{b} + (4\,\lambda_2 / n)\;\mathbf{d}^*\,\mathbf{b} - 6\,\lambda_3\,\|\boldsymbol{\xi}\|_p^{\,2-p}\,\mathbf{w}$$

  where **w** is a matrix with elements

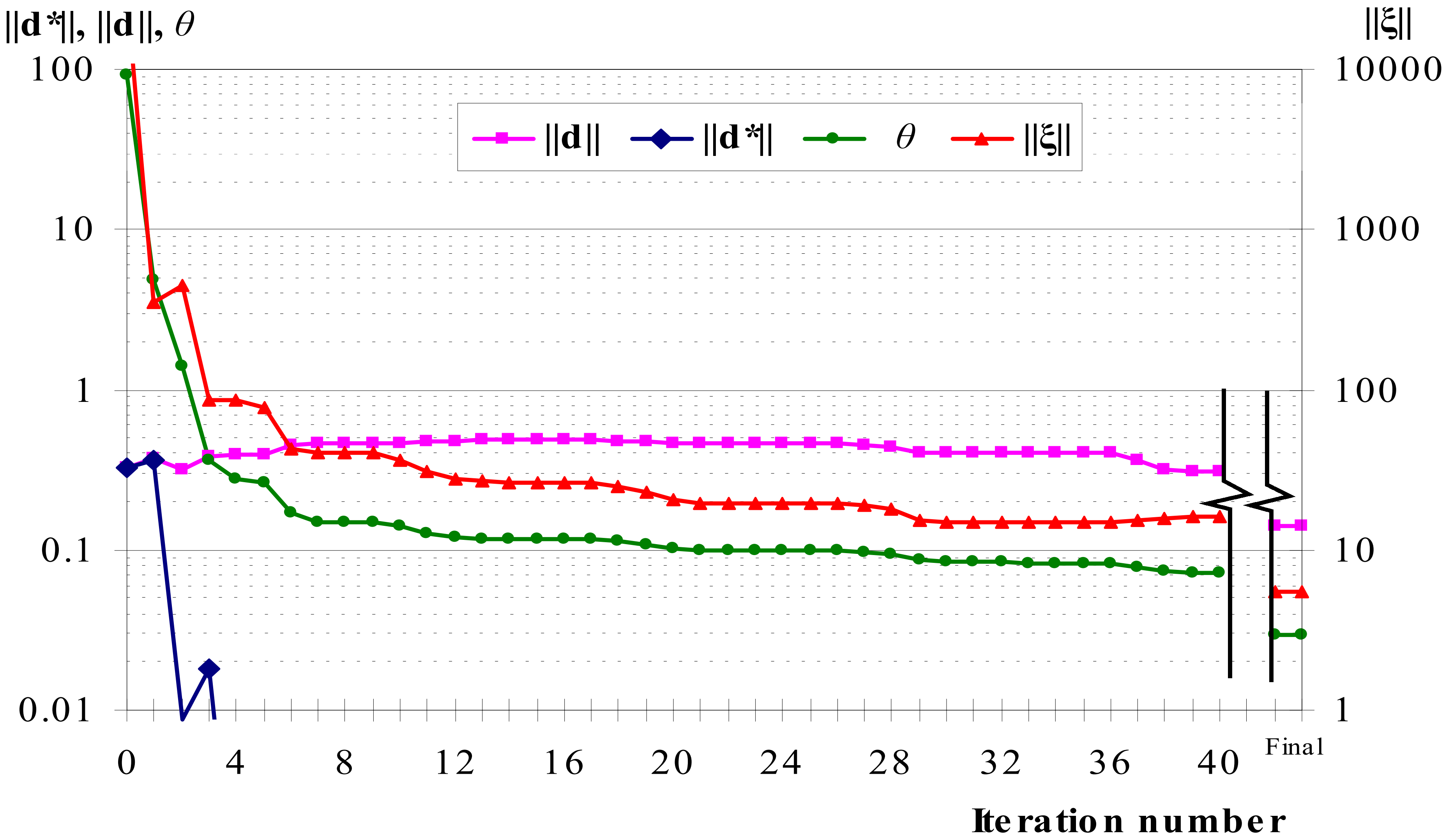
  $$w_{ij} := \xi_j\,\psi_i$$

  and **ψ** is a vector defined by

  $$\boldsymbol{\psi} := \{[\mathbf{b}'^{(3)}]^{-1}\}^T\,\boldsymbol{\xi}^{(p-1)}$$

- This enables the use of typical nonlinear optimisation methods such as the Fletcher-Reeves Conjugate Gradient method
- The initial value of **b** could be either the Cholesky solution or even the identity matrix

Evolution of solution through iterations – Example B



A note on disaggregation problems

- The proposed technique is directly applicable to disaggregation models
- All-at-once disaggregation models such as Schaake-Valencia or Mejia-Rousselle may involve huge sizes of matrices with an unreasonably high number of parameters
- The proposed technique is strongly recommended for coupling with the *Simple Disaggregation* model (Koutsoyiannis and Manetas, *Water Resources Research*, 32(7), 1996) whose parameters coincide with those of the typical multivariate PAR(1) model

Conclusions

- The problem of preserving skewness in stochastic hydrologic models is directly associated to the problem of covariance matrix decomposition
- A new technique is presented for covariance matrix decomposition based on an optimisation framework, with the objective function being composed of three components aiming at
  - complete preservation of the variances of variables
  - either preservation of covariances, or optimal approximation thereof (in case of inconsistent covariance matrices)
  - preservation of the skewness coefficients by keeping the skewness of the noise variables as low as possible
- The technique is implemented by a simple nonlinear optimisation algorithm based on analytically determined derivatives
- Applications indicate that the algorithm is quick, stable and easily applicable even in cases with as much as 1600 parameters